



# National Plant Genome Initiative

## Progress Report January 2004

**National Science and Technology Council**  
**Committee on Science**  
*Interagency Working Group on Plant Genomes*



Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>JAN 2004</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2004 to 00-00-2004</b>	
4. TITLE AND SUBTITLE <b>National Plant Genome Initiative. Progress Report</b>			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>National Science and Technology Council, Executive Office of the President, 725 17th Street Room 5228, Washington, DC, 20502</b>			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES <b>32</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

## About the National Science and Technology Council

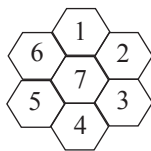
The National Science and Technology Council (NSTC), a cabinet-level Council, is the principal means for the President to coordinate science, space, and technology to coordinate the diverse parts of the Federal research and development enterprise. An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems and strengthening fundamental research. The Council prepares research and development strategies that are coordinated across Federal agencies to form an investment package aimed at accomplishing multiple national goals.

Additional information regarding the NSTC can be obtained from the NSTC website, [http://www.ostp.gov/NSTC/html/NSTC\\_Home.html](http://www.ostp.gov/NSTC/html/NSTC_Home.html).

**Note:** This document does not represent the final determination in an overall Administration budget decision-making process. The programs presented in this report will have to compete for resources against many other high-priority Federal programs. If these programs compete successfully, they will be reflected in future Administration budgets.

### Cover Photo:

Photos 1 through 6 courtesy of the Grain Inspection, Packers and Stockyards Administration, USDA; Photo 7 courtesy of the Agricultural Research Service, USDA



- |            |         |
|------------|---------|
| 1. Barley  | 5. Rye  |
| 2. Oats    | 6. Corn |
| 3. Wheat   | 7. Rice |
| 4. Sorghum |         |

Cover Design by Victoria Bryan of the National Science Foundation.

# National Plant Genome Initiative

## Progress Report January 2004

National Science and Technology Council

*Committee on Science*

*Interagency Working Group on Plant Genomes*





EXECUTIVE OFFICE OF THE PRESIDENT  
**NATIONAL SCIENCE AND TECHNOLOGY COUNCIL**  
WASHINGTON, D.C. 20502

Dear Colleague:

This report provides an update on the National Plant Genome Initiative (NPGI). The report was prepared by the National Science and Technology Council's Interagency Working Group on Plant Genomes (IWG), which coordinates and provides oversight for the Federal investment in plant genome research. As part of its responsibility, the IWG monitors the progress of the NPGI and documents significant progress in annual reports.

In January 2003, the IWG published a new 5 year plan, "National Plant Genome Initiative: 2003-2008," which outlined six broad objectives for the NPGI. Significant progress has been demonstrated on all six objectives during the past year as illustrated by examples highlighted in this annual report. Tools for plant genomics research are being developed, and a trend toward increased activities in functional genomics and translational genomics is beginning to emerge. A hallmark of the NPGI is open and free sharing of research results, allowing a broad community of scientists to participate in plant genomics research. There is no doubt that the NPGI is making a major contribution to science and is having an impact on society through the research community's efforts in education, training, and outreach.

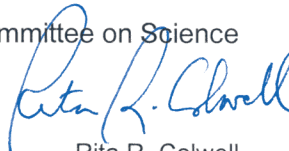
Judging from the accomplishments described in this report, there is every indication that significant advances will be made in the coming year. The IWG will continue to coordinate the NPGI to ensure that U.S. efforts in plant genomics benefit from interagency support and cooperation, keeping U.S. scientists at the forefront of plant biology and its application to solving global problems in agriculture, public health, energy, and environmental protection.

Sincerely,

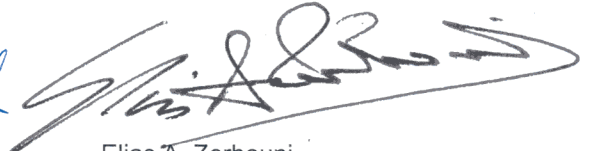
Committee on Science



Kathie L. Olsen  
Associate Director for Science  
Office of Science and Technology Policy



Rita R. Colwell  
Director  
National Science Foundation



Elias A. Zerhouni  
Director  
National Institutes of Health

---

# **Interagency Working Group on Plant Genomes Committee on Science National Science and Technology Council**

## Co-Chairs

Mary E. Clutter  
Assistant Director  
Biological Sciences  
National Science Foundation

Joseph Jen  
Under Secretary  
Research, Education, and Economics  
U.S. Department of Agriculture

## Members

Robert B. Bertram  
Agricultural Research Specialist  
Bureau for Economic Growth,  
Agriculture and Trade  
U.S. Agency for International  
Development

Noah Engelberg  
Resource Management Officer  
Office of Management and Budget

Brad Fenwick  
Chief Science Advisor  
Cooperative State, Research,  
Education and Extension Service  
U.S. Department of Agriculture

Clifford Gabriel  
Deputy to the Associate Director  
Office of Science and Technology  
Policy

Jane L. Peterson  
Associate Director  
Division of Extramural Research  
National Human Genome Research  
Institute

Judy St. John  
Associate Deputy Administrator  
Agricultural Research Service  
U.S. Department of Agriculture

Walter Stevens  
Division Director  
Division of Chemical Sciences,  
Geosciences, and Biosciences  
Office of Science  
U.S. Department of Energy

## Executive Secretary

Machi F. Dilworth  
Division Director  
Division of Biological Infrastructure  
National Science Foundation

## Table of Contents

---

ExecutiveSummary.....	1
Introduction .....	3
Progress Reported in the Past Year.....	5
New Projects Started in 2003.....	16
Plans for the Next Year.....	21
Appendix.....	23

## *I. Executive Summary*

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research program by the Interagency Working Group (IWG) on Plant Genomes with representatives from the Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), and Office of Management and Budget (OMB). In 2003, the Agency for International Development (USAID) joined the IWG. Since 1998, the IWG has provided overall coordination and oversight of plant genome research activities supported by the NPGI participating agencies, as outlined in the initial five-year plan published in January 1998 (<http://www.ostp.gov/NSTC/html/npgireport.html>).

In January 2003, the IWG published a new long-range plan (<http://www.ostp.gov/NSTC/html/npgi2003/index.htm>). The plan incorporated recommendations from a broad community of stakeholders, including a report from the National Research Council, as well as IWG's own assessment of the first five years of the NPGI. By all accounts, the first five years of the NPGI were a resounding success. Federal investments in plant genome research since 1998 have galvanized the plant research community and helped place the US at the forefront of plant genomics in the world. The new plan has outlined six broad objectives designed to continue and accelerate advances in plant genomics: (1) Continued elucidation of genome structure and organization; (2) Functional genomics; (3) Translational plant genomics; (4) Bioinformatics; (5) Education, training and outreach; and (6) Broader impacts.

In this report, continued progress in plant genomics research is described by providing illustrative examples of research results reported during the past year. They include:

- Construction of a high resolution maize map that integrates genetic and physical maps - a culmination of five years of hard work - which will benefit both basic researchers and breeders
- Identification of the full encyclopedia of genes necessary for mineral nutrition in plants, that forms the foundation for understanding the mechanism of plant uptake of both beneficial and toxic minerals
- Development of the marker-assisted breeding strategies for wheat
- Establishment of a comparative cereal genomics database, Gramene, which uses the complete rice genome sequence as a reference and serves as the information resource for the entire cereal research community including maize, wheat, barley and sorghum
- Active involvement of plant genome researchers in education and training of undergraduates, high school students and K-12 teachers, which is contributing to an increased number of US students interested in studying plant sciences
- Research collaboration between US scientists and scientists in developing countries in plant genomics and related fields of science

Also reported are some examples of new projects that promise to advance the field in the coming years. They include:

- Building of resources and tools for plant genome research, such as a set of enzymes to be used as a tool to study plant cell walls, community microarray facilities for rice and maize, and a rice genetic stock center

## *Executive Summary*

---

- Nutritional genomics of sunflower with the ultimate goal of improving the quality of sunflower oil
- Identification of a network of genes involved in disease resistance in soybeans
- A new comprehensive database for the entire plant genome research community to provide seamless access to relevant information resources that are distributed all over the world

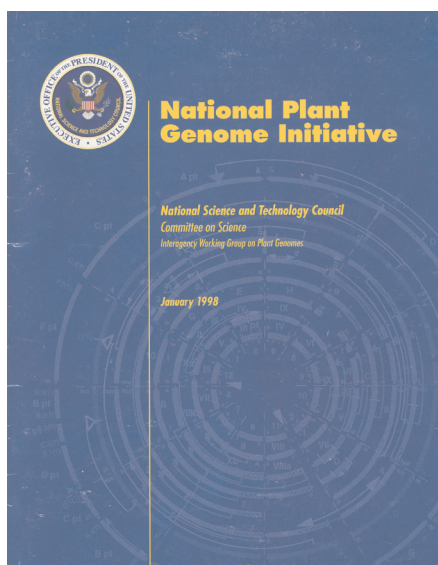
In the coming year, all agencies participating in the NPGI plan to continue support of plant genome research based on the five-year plan published in 2003. At the same time, the IWG and NPGI participating agencies are well aware of the rapid pace of scientific and technical advances in a field like genomics, and remain ready to support new opportunities as they arise. The IWG will continue to coordinate and provide oversight to the NPGI.

## II. Introduction

2003 marked the 50<sup>th</sup> anniversary of the discovery of the structure of DNA by Watson and Crick. This discovery changed the course of biology forever, ushering the world into the age of molecular biology, genetic engineering, and now genomics. These advances in our fundamental understanding of nature, and technological innovations have impacted our way of thinking about life on earth, and there is no sign that the biology revolution will stop at genomics.

Plant biology has been transformed completely over the last 50 years. It is now squarely in the age of genomics, and is constantly changing as new concepts emerge and novel technologies develop. Recognizing the enormous scientific opportunities, the National Plant Genome Initiative (NPGI) was established in 1997 under the National Science and Technology Council (NSTC) and the Office of Science Technology and Policy (OSTP). The NPGI is coordinated by the Interagency Working Group on Plant Genomes (IWG). The IWG is charged with identifying science-based priorities for a national plant genome initiative and to plan and coordinate Federally supported genome research activities for the Nation.

The IWG was established in May 1997 by the Office of Science and Technology Policy in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee. The IWG currently consists of representatives from National Science Foundation (NSF), Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), Agency for International Development (USAID), Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB).



In January 1998, the IWG published the first NPGI Five-Year Plan (<http://www.ostp.gov/NSTC/html/npgireport.html>). The plan outlined the goals of the NPGI, objectives for 1998-2002, and guiding principles for a successful national program. Subsequently, the IWG published annual progress reports and documented the accomplishments against the goals and objectives outlined in the plan. By all accounts, the first five years of the NPGI have been a success. Increased Federal investments in the NPGI contributed to advances in plant genomics research, enabling the US to be at the forefront of plant genomics. Assessment of the first five years of the NPGI led to a conclusion that; “it is critical to continue and even accelerate research efforts in plant genomics in order to take advantage of exciting scientific opportunities that will lead to improved agriculture, energy and health, thus ensuring a high quality of life for future generations” (a statement by Dr. John H. Marburger, Director of the Office of Science and Technology Policy, excerpted from the transmittal letter attached to the new NPGI five-year plan). Accordingly, the IWG published the second NPGI Five-Year Plan

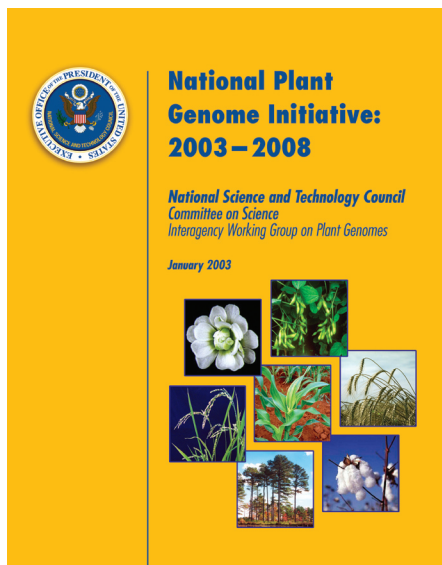
## Introduction

---

(<http://www.ostp.gov/NSTC/html/npgi2003/index.htm>). The plan summarizes major accomplishments since 1998, and outlines new objectives for 2003-2008 while reaffirming the NPGI goals and guiding principles as established in the first five-year plan. Six major objectives are listed in the report:

- Continued Elucidation of Genome Structure and Organization
- Functional Genomics
- Translational Plant Genomics
- Bioinformatics
- Education, Training and Outreach
- Consideration of Broader Impacts

What follows are examples of accomplishments reported since January 2003, as an illustration of continued advances in plant genomics research. The IWG is pleased to note that progress has been reported in all of the six objectives, and there is every reason to be optimistic about expanding and continued advances in plant genomics in the US.

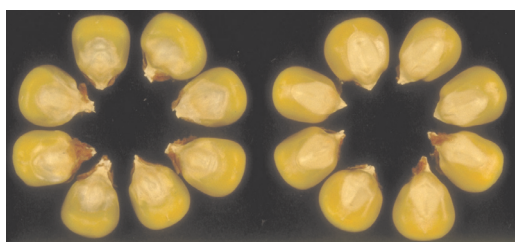




### ***III. Progress Reported in the Past Year***

Many of the projects funded in early days of the NPGI are beginning to produce tangible results in the form of new information, deeper understanding of the biology of plants, and useful research tools, contributing to advances in plant biology and genomics.

- ***Continued Elucidation of Genome Structure and Organization***



#### High density Maize physical map produced

The goal of one of the first NPGI projects funded, the Maize Mapping Project led by the University of Missouri, was to develop a richly detailed and well-integrated physical and genetic map of the maize genome. This year, the researchers announced that they are entering the home stretch. More than 95% of the physical map has been assembled and much of it linked to the genetic map that was made available in November 2002. Even as it enters the final phases of

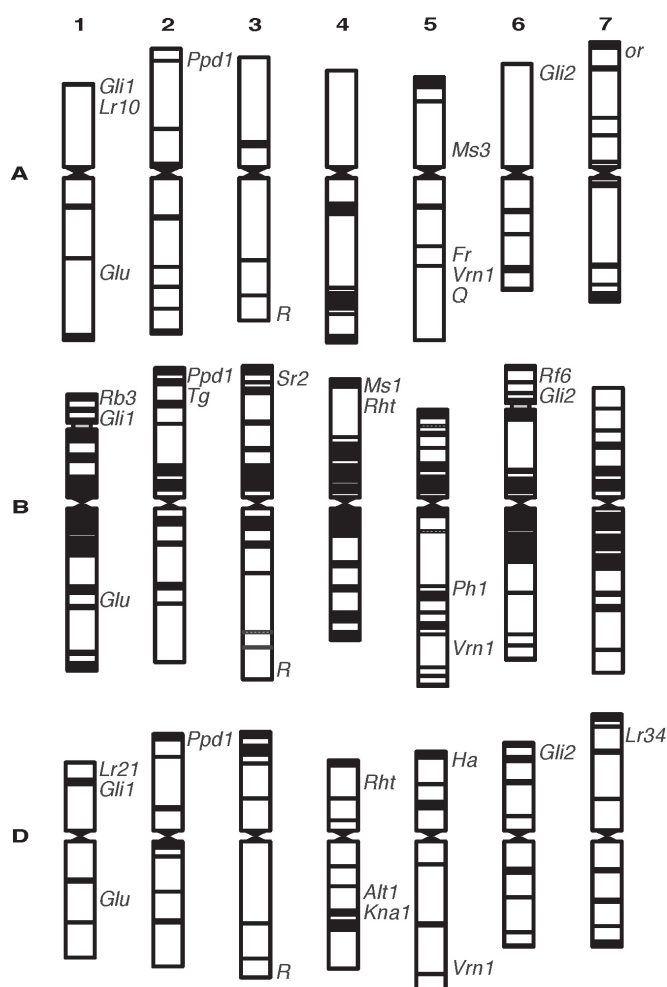
completion, it has already had a tremendous impact on maize researchers because it allows location of genes based on their map positions. The current map can be accessed at <http://www.maizegdb.org/> and details of the mapping project can be found at <http://www.maizemap.org/>. This map will benefit the research and breeding communities in many ways. For example, it provides an extraordinarily high-resolution map to corn breeders as well as researchers doing fundamental genetics. This map will be an essential resource for efforts to sequence the gene rich regions of the maize genome that will likely ramp up over the next 1-2 years.

#### Rice chromosome 10 completely sequenced and analyzed

Recently, the US Rice Genome Sequencing Project and collaborators completed the sequencing and analysis of the 22-megabase rice chromosome 10. ("Rice Chromosome 10 Sequencing Consortium", *Science* 300:1566-1569, 2003). Although chromosome 10 is the smallest rice chromosome, the gene content was almost two times higher than expected, with approximately 3500 genes identified. There was also a high degree of similarity between the coding regions of this portion of the rice genome with the other completely sequenced plant genome, *Arabidopsis*. However, unlike *Arabidopsis*, rice chromosome 10 displays little evidence of recent large-scale sequence duplications.

As expected from the draft sequence, colinearity was evident with other cereals such as sorghum and maize. Thus, the benefits of sequencing the rice genome are not limited to rice and will be seen in other crop species of significant economic importance in the US such as wheat, barley, corn and sorghum. These results reiterate the importance of completing this second plant reference species genome to fully exploit the biological diversity between monocotyledonous (e.g. grasses) and dicotyledonous (e.g. soy, cotton, and *Arabidopsis*) flowering plants. With the completion of chromosome 10, the US Rice Genome Sequencing Project members (University of Arizona, Clemson University, Cold Spring Harbor Laboratory, Washington University in St. Louis, and The Institute

for Genomic Research) are now focused on finishing chromosomes 3 and 11 as well as other regions to fulfill the International Rice Genome Sequencing Project commitment of finishing the entire rice genome by the end of 2004.



### Evolution of wheat genome structure and organization

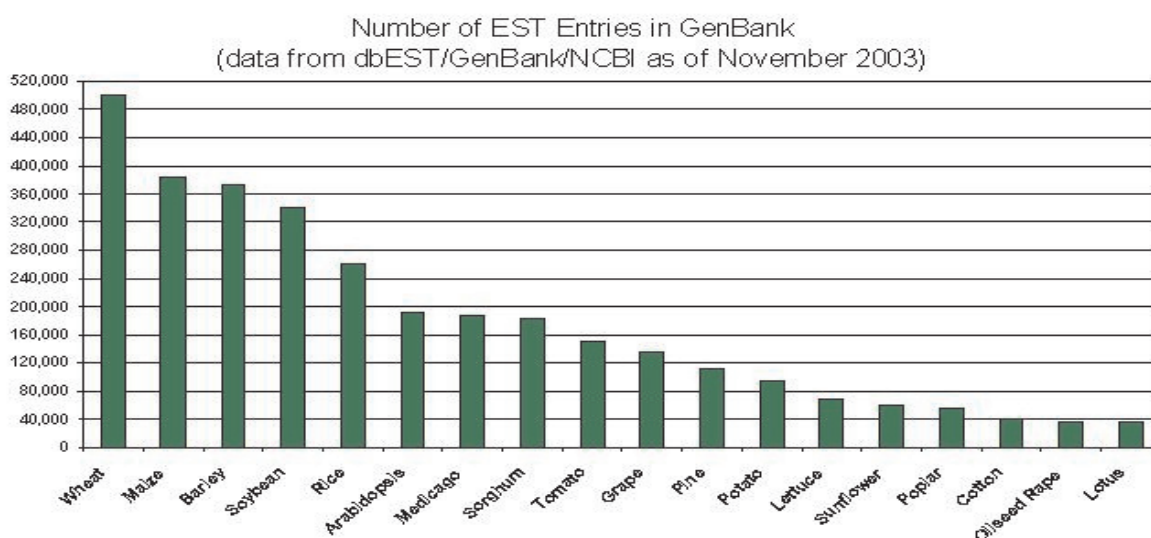
A project led by the University of California, Davis, has been genetically mapping Expressed Sequence Tags (ESTs) to the bread wheat genome. Since each EST represents a piece of sequence from an expressed gene, this mapping effort gives a snapshot of the distribution of the genes throughout the genome. The genome of bread wheat is hexaploid (6N) and actually resulted from the combination of genomes from three different diploid (2N) wheat varieties. The bread wheat genome has undergone a tremendous amount of change since the three genomes came together, with evidence of gene duplication, deletion, and recombination, or swapping pieces of related genes. Which of these enormous genome changes has brought the desirable agronomic traits that make bread wheat a staple grain through much of the world is not yet known.

The EST mapping efforts have already provided some clues about how these forces are shaping the genome. For instance, genes present in a single copy tend to be located near the centromere, where relatively little recombination takes place, while duplicated genes are found near the ends of the

chromosomes where recombination is high. Approximately a quarter of the genes examined were members of duplicated gene sets. In cases where both the ancestral gene and its duplicate copies were mapped, the duplicate 'offspring' were closer to the chromosome ends than the ancestral copy. These observations create a picture of a dynamic genome, where higher recombination rates occurring at the chromosome ends allow more rapid rates of evolution in the duplicated genes. This would allow the plant to try out 'new versions' of genes while maintaining the original functions elsewhere in the chromosome. Such studies allow us to gain insights into the evolution of agriculturally important plants and may lead to discovery of genes that are important in domestication and conferring desirable traits.

More than 3 million plant EST sequences in the public database

The number of ESTs (Expressed Sequence Tags) in the dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST/>) continues to increase at a rapid rate. At the beginning of the NPGI, the number was approximately 175,000. As of November 2003, that number is over 3 million, and researchers are using the EST sequence information in dbEST in an increasing number of ways. They are especially useful for cataloguing genes in related plants, and as a mapping tool. Wheat has provided the most dramatic increase in the number of entries over the past year. This has come about from the combination of deposits from the publicly supported wheat genome EST project and the release of data by private industry.



- **Functional Genomics**

Genomes of the two reference plant species have been sequenced, and researchers are using the primary sequence information to understand the biological role of genes, regulatory elements and repeated sequences.

Essential tools for functional genomics in *Arabidopsis*

A major breakthrough in plant functional genomics was reported this year for *Arabidopsis thaliana* in a project led by the Salk Institute (“Genome-wide insertional mutagenesis of *Arabidopsis thaliana*”, *Science* 301:653-657, 2003). 225,000 unique insertion events were created using *Agrobacterium* T-DNA. The exact locations of 88,000 of these insertions were identified by DNA sequencing. This enormous effort created a unique and extremely valuable resource for the community: a ‘sequence indexed’ collection of insertion mutants in 21,700 of the 29,000 predicted genes of *Arabidopsis*. The

entire collection is made freely available to the community from the *Arabidopsis* Biological Resource Center at Ohio State University.



These lines allow genome-wide functional analysis of genes or gene families of interest, and should revolutionize the way in which plant functional genomics is conducted. In a proof of concept experiment, the authors combined microarray and bioinformatic analyses to identify four hormone inducible genes that encode two domains characteristic of plant regulators of transcription. Deletion mutants in each of these specific genes were tested, and it was found that these are likely to be transcription factors necessary for

normal hormone responses. These gene functions would not have been discovered in classical genetic screens: the genes have overlapping functions, and multiple mutant plants were required to observe the hormone response defects.

### Mineral Nutrition in Plants

Understanding how plants take up minerals from the soil allow the development of new strategies for fortification of plant foods with nutrients beneficial to human and animal health (iron and calcium, for example) and might suggest new approaches for using plants to clean up soils with unsafe levels of heavy metal pollutants such as cadmium, lead and mercury. The systems that allow uptake of beneficial and toxic ions in plants are extremely complex, and as a result are generally not well understood at the molecular level.

A project led by Dartmouth University is taking an ambitious functional genomics approach to identifying the full encyclopedia of genes needed for mineral nutrition in plants. Using a method called inductively coupled plasma spectroscopy (ICP-MS), the team has screened 6,000 *Arabidopsis* mutants simultaneously for changes in 18 elements (“Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*”, *Nature Biotechnology* 21:1215-1222, 2003). This very successful screen yielded 50 new mutants, including one in a gene known to be involved in uptake of iron, a key nutrient for human health, thus validating the approach.

Based upon past studies of genomic and EST sequences, it was predicted that plants have hundreds of genes needed for importing minerals and heavy metals. This makes it very difficult to use bioinformatics to accurately predict the genes that play indispensable roles in plant nutrition. The results of this functional genomics screen agree with this estimate, and show that functional genomics approaches will be needed to unravel the “ionome”. These studies establish ICP-MS screening as an efficient method for nutritional functional genomics in *Arabidopsis*, and the results of these experiments will suggest approaches for improvement of crop plant nutritional value.

### Potato late blight disease resistance genes



The potato genome project was started in 1999 to develop tools for potato functional genomics. At its heart was the goal of developing tools to allow researchers and breeders to pinpoint the genes that make wild potato varieties such as *Solanum bulbocastanum* much more resistant to disease than their cultivated counterpart *Solanum tuberosum*. There is an urgent need for these tools because *Phytophthora infestans* (late potato blight), the causative agent of the Irish potato famine of the 1840s, is once again posing an increased threat to agriculture. This oomycete or

‘water-mold’ disease can effectively destroy an entire potato field since none of the commercially-grown varieties can completely resist infection.

The project has made two significant breakthroughs in the past year. First, a broad-spectrum late potato blight resistance gene, called *RB*, has been isolated from *S. bulbocastanum*. The *RB* gene is unlike any previously isolated *P. infestans* resistance genes because it confers resistance to all strains of *P. infestans*, not just one or two. Second, the project team has been able to determine the structure and organization of two large potato disease resistance gene clusters. The two regions include more than 30 resistance gene candidates and at least three active resistance genes that confer resistance to disease-causing viruses and *P. infestans*, as well as genes for broad spectrum pathogen resistance. These genome regions are now being sequenced and the resulting information will provide a treasure trove of candidate genes for developing potato varieties that can withstand the challenges of destructive pathogens such as *P. infestans* as well as new pathogens that might emerge in the future.

- ***Translational Plant Genomics***

Fundamental understanding of the gene structure and function is being applied to improve the quality of economically important plants.

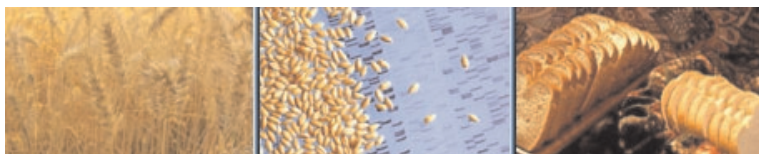
### Tailoring corn gene expression using RNAi

The maize Opaque2 transcriptional activator plays an important role in multiple seed-specific metabolic pathways; however, previous efforts to exploit its regulatory properties to improve amino acid content in corn were hampered by unintended effects on other genes controlling tissue texture and pest resistance. Plant genome investigators at Rutgers University recently applied RNAi approaches, initially developed in *Arabidopsis*, to generate transgenic maize that exhibited a dominant Opaque2 phenotype, thus achieving success in translational genomics. These plants selectively turn off production of less desirable zein storage proteins while retaining expression of other members of this large multigene family and other agronomically desirable characteristics. As one of the first



reports of stable alteration of a cereal seed component by RNAi, future applications may include tailoring plants for improved nutrient or energy storage qualities.

### Marker assisted wheat breeding



Recent progress in plant genomics has the potential to initiate a new “Green Revolution”. To bring this potential to the grower’s fields, these discoveries need to be incorporated into commercial

varieties. Recent progress in molecular genetics has resulted in the development of DNA tags, which can be deployed in Marker-Assisted Selection (MAS) strategies for cultivar development. These molecular markers can be used as chromosome landmarks to make the selection of useful agronomic traits easier. This technique is particularly useful for: (1) genes that are highly influenced by the environment; and (2) genes for resistance to diseases for which screening is difficult. They are also useful to accumulate multiple genes for resistance to specific pathogens and pests within the same cultivar, a process called gene pyramiding.

Because wheat is a self-pollinating species, growers can save part of the grain from one harvest to use as seed the next year, thus limiting profitability of wheat breeding and reducing private sector involvement. As a consequence, approximately 60% of the wheat cultivars released in the USA during the 20<sup>th</sup> century have been developed through publicly funded breeding programs. At the end of 2001 a national wheat MAS consortium was begun, which includes wheat molecular geneticists and breeders from 12 public programs across the US. The overall goal of the wheat MAS (MASwheat) project is to efficiently transfer genes encoding useful traits into 75 cultivars and wheat lines adapted to the main production areas of the US. Target traits include: resistance against fungi, viruses, and insect pest; and bread, pasta, and noodle quality. Eight generations will be required to complete the transfer of the identified genes. Two generations of crosses are advancing per year, resulting in a total of 1,050 marker-assisted crosses in the first three generations. All the information generated by MASwheat is publicly available at <http://MASwheat.ucdavis.edu/>. MASwheat activities so far have resulted in seven scientific publications and 23 presentations in growers meetings, field days and symposia aimed at improving public understanding of the benefits of biotechnology. MASwheat has also created an integrated network of breeders and researchers across the country facilitating the transfer of knowledge and germplasm.

### Public seed initiative

In recent years an explosion of knowledge about crop genomes has resulted in the identification of many genes responsible for important crop traits. This is especially true in species such as rice and tomato, which are well suited for genetic studies. A multidisciplinary team of plant breeders, molecular biologists, USDA and extension personnel and non-profit groups has partnered to build on these investments and enhance the



delivery of publicly bred vegetable varieties through the Public Seed Initiative (PSI; <http://www.plbr.cornell.edu/psi>).

The PSI aims to extend from the genomics lab to improve the delivery of benefits to farmers and consumers. Existing grower networks in the Northeast and Northwest have been recruited to conduct on-farm trials of new varieties developed with tools from genomics research. Links between public breeders and seed companies, large and small, have also been strengthened.

Through the PSI, more than a dozen public varieties are being evaluated through extension networks, by companies, and on farms from Maine to California, and viewed by wide audiences at a series of annual field days. Hundreds of growers have attended seed production workshops and hands-on plant breeding workshops. Based in part on demand created by participatory trials, a number of these plant varieties and breeding lines have been licensed on a non-exclusive basis to recipients including large multinational seed companies, smaller companies focused on regional, organic, or specialty markets, and distributed to non-profit groups interested in genetic diversity and sustainable agriculture. Results from these trials have also identified new objectives for vegetable breeding programs, expedited by knowledge and tools from crop genomics and farmer demand.

- ***Bioinformatics***

Plant genome bioinformatics efforts are moving away from project-specific databases serving a small group of scientists to larger databases housing data for a particular organism or a particular process. Accordingly, community engagement has become a central part of database design. The three examples of community databases highlighted below both represent a trend towards engaging the end-users in development of the resource itself.

MaizeGDB – An integrated data resource for Maize

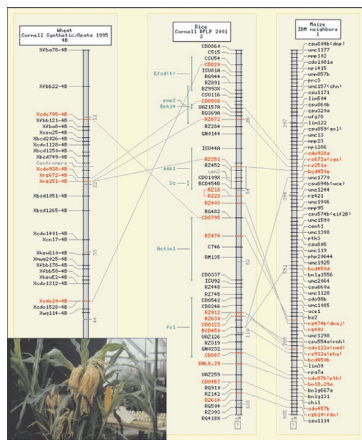
The burgeoning resources for maize genomics have done more than impact researchers at the bench. In the past two years, there has been a quiet revolution in the databases that for the last decade housed maize genetics data and served the broader community. The data in MaizeDB, a USDA-ARS database, and ZmDB, the database housing EST and genome survey sequence from The Maize Gene Discovery Project, have been merged to form “MaizeGDB” (<http://www.maizegdb.org/>). This new resource was developed to house not just the data in hand but future types of data. The database structure is modular so that different data types can be housed and viewed in a flexible way. The resources are broadly accessible to students, new researchers interested in maize biology, and breeders (<http://www.maizegdb.org/education.php>). MaizeGDB is rapidly becoming a one-stop shopping location for maize genome resources and should grow and evolve with the data and community needs.

Gramene – A comparative cereal genomics database

An international collaboration that includes US laboratories is on the verge of finishing the rice genome; within the next year the complete compendium of rice genes will be available to the world. Rice is a key staple food, which together with wheat, barley, corn and sorghum forms mankind’s most important source of calories. It is also an increasingly important model for cereal functional



genomics. Despite the dramatic difference in the shape, size and growth habits of the cereals, they are all highly related and shared a common ancestor approximately 50 million years ago. Because of their interrelatedness, rice, wheat, corn, sorghum and the other cereals share most of the same genes, and have similar biology. But the rice genome, at 430 million base pairs, is much smaller than the genomes of the other cereals, which are 5 to 100 times larger. By studying the rice genome, researchers hope to understand the genomes of the other cereals, and by doing so to make discoveries that will increase their yield, taste, disease resistance, nutritional value and other desirable traits.



In a project led by the Cold Spring Harbor Laboratory in NY, researchers have created “Gramene,” an online database for comparative cereal genomics. This web-accessible database (<http://www.gramene.org/>) represents the genome of rice (for two different sequenced varieties) and the genome’s known and predicted genes. For each gene, Gramene provides researchers with information on its location in the genome, its putative function, similarities to genes in other plants and animals, and information on variants of the gene. A key feature of Gramene is its use of comparative genetic maps. Using a variety of bioinformatics techniques, the Gramene project has aligned the maps of wheat, barley, corn, sorghum and oats to the rice genome and makes the comparative maps available for searching and browsing. These maps make it possible for a researcher who is studying a trait that has been genetically mapped

on wheat and corn or another cereal to quickly find the corresponding region in the rice genome. Once the corresponding rice region is known, researchers can “zoom in” to find what rice genes are present in that region, download their DNA and protein sequences, and view information about mutations or other variants in the gene that affect the growth or development of the plant. This information is invaluable to researchers who are searching for the genes responsible for a desirable trait.

Gramene is a step towards a comprehensive database of cereal genomes. The next step will be to incorporate other cereal genomes into Gramene. As these genomes become available, their DNA sequences will be aligned to rice, providing researchers with increasing detailed and comprehensive maps of the relationships among the grain genomes. These resources will greatly accelerate the ability of plant biologist to understand the unique biology of the cereals, and for breeders to develop new varieties that are naturally resistant to diseases, drought and other environmental stresses, or that have increased nutritional value.

### PlantsP

PlantsP (<http://plantsp.sdsc.edu/>) is a plant kingdom-wide database that aims to capture and make accessible all that is known about protein phosphorylation and dephosphorylation in plants.

Why dedicate a database to these pathways? Protein phosphorylation and dephosphorylation are fundamental to regulation of a multitude of cellular processes in plants, animals and microbes. These post-translational modifications, which are catalyzed by protein kinases and phosphatases, together form a reversible molecular switch. In plants, this switch has been implicated in the control of

most of the major developmental events and environmental responses including cell cycle control, transcriptional and translational regulation, control of carbon and nitrogen metabolism, regulation of growth and differentiation, and responses to abiotic and biotic environmental cues. The ubiquitous use of the phosphorylation/dephosphorylation switch is reflected in the number of genes encoding enzymes that perform these tasks. For example, *Arabidopsis* has approximately 1000 kinase genes, 300 phosphatase genes, and at least 50 genes involved in regulation of these reactions, together comprising 1/20 of the genome. Because protein kinases and phosphatases control so many processes in plants, a comprehensive database is essential to understand the interrelationship all of the networks they regulate. PlantsP provides up-to-date cross-searchable catalogs of all known genes and associated resources for their study, and includes opportunities to comment on and participate in the annotation of these important genes.

- ***Education, Training and Outreach***

Plant genomics research represents the cutting edge of today's biological research. Instead of studying one gene at a time, researchers study a network of genes or even the whole genome at once. Genomics is inherently multi-disciplinary and uses concepts and technologies developed not only in biology, but also computer sciences, chemistry, mathematics, and engineering. As such, plant genomics research provides an exciting opportunity for students to be exposed to a new way of doing science. Training of the next generation of scientists is one of the most important goals set for the NPGI.

Campus-wide education and outreach programs

Various institutions have developed larger training and outreach programs that capitalize on the breadth and depth of opportunities available through genomics research.

The University of California, Davis has recognized the value of outreach activities initiated by projects supported through the agencies participating in the NPGI and has contributed additional support. The UC Davis Partnership for Genomics Education (<http://ceprap.ucdavis.edu/>) is dedicated to developing an educational program focused on plant genomics and biotechnology targeted towards secondary level students. Several very successful outcomes of this program are the development and dissemination of educational software and on-line materials, development of associated hands-on activities, equipment loan programs, teacher training, and student internships. The educational software currently available includes a popular virtual DNA fingerprinting laboratory and will soon include a virtual plant genomics laboratory.

A vibrant training environment was established at the Boyce Thompson Institute for Plant Research and Cornell University in Ithaca, NY for young researchers interesting in participating in cutting-edge genomics research in plants. The Emerson Summer Genetics Program (ESGP) (<http://outreach-pgrp.cornell.edu/index.htm>) provides an opportunity for undergraduate students, high school students and high school teachers to participate in development of transposon-tagged maize lines for the research community. This highly successful program attracted 8 students for summer field and lab work in 2002. The students



ESGP 2002 Class



ESGP 2003 Class

carried out pollinations in the field and assisted in maintaining the corn population. In the laboratory, they performed analyses of the tagged lines.

A common observation is that once plant genome awardees establish successful training programs, their colleagues join to create an even more successful and larger outreach endeavor. This program was no exception: in 2003 several other local laboratories became involved in this program and trained a total of 17 students. Participants included 10 students from area high schools in Ithaca and surrounding areas, 5 of whom were from underrepresented minority groups. A full-time coordinator was hired this year and is providing cohesion for all the plant genome project's outreach efforts at Cornell University and the Boyce Thompson Institute.

#### Teacher training at the University of Georgia at Tifton

A variety of teacher training projects continue to serve the needs of local communities, using Plant Genomics as a focus. A project led by the University of Georgia, Tifton, targets teachers from rural areas of Georgia, where low population density and distance from the universities in Athens and Atlanta create a need for training teachers to bring back the latest scientific advances to their home institutions and act as agents of change. In this program teachers participate in a four-week course with several complementary components. First, they are exposed to available web resources on Georgia science curriculum as well as US-wide educational web sites. They then share their teaching approaches with each other and the participating scientists. This provides the participants with a chance to discuss their experiences and bring new tools back to their home institutions. Next they spend three weeks in the PI's laboratory learning DNA methodologies and getting experience in bioinformatics. At a later date the student teachers participate in a Regional Education Service Agency 'In Service' Workshop from which they bring back an electrophoresis 'discovery kit' for use in their home school.

#### • ***Broader Impact Issues***

Science has become truly global and barriers of all kinds are disappearing. In order for the NPGI to be at the forefront of plant genomics, it is important that the research community forges collaboration with international partners and with the private sector. US plant genome researchers have been actively

seeking partnerships with their colleagues in other countries and outside the academia. They are especially concerned about the lack of access to the newest research by their colleagues in developing countries, and some have started such collaborations.

#### Awardee activities supporting agriculture in developing countries

A project at Texas A&M University is using the Sorghum genome map to tease out the networks of genes that control drought tolerance. Sorghum, a grass that originated in Africa, is the fifth most important cereal worldwide. It has evolved characteristics, such as thick waxy leaves and a deep root system, that allow it to grow in hot, dry climates where water is limited. The sorghum genome is similar to those from other important cereals such as rice, corn, and wheat. This research will have an impact on a crop that is widely grown in Africa and India. Several African scientists will work with the PI in the US, and she will also travel to Africa to participate in workshops and training activities.

Scientists at the University of Georgia are collaborating with researchers at ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) to develop more useful simple sequence repeat (SSR) genetic markers for sorghum. The SSR markers will be applied to studies documenting germplasm diversity and to marker-assisted breeding programs for agronomic targets important in sub-Saharan Africa. The collaboration brings together researchers at ICRISAT, who have significant experience collecting and documenting wild sorghum and millet species from the entire African continent, with cutting edge plant genomics researchers in the US.

A project led by a scientist at the University of Virginia will focus on mechanisms for resistance of African cowpea to the parasitic plant *Striga* (witchweed: <http://pi.cdfa.ca.gov/weedinfo/STRIGA2.html>). Cowpea, an annual legume, originated in Africa and is widely grown in Africa, Latin America, Southeast Asia and in the southern United States. It is chiefly grown for animal fodder, or as a vegetable for human consumption. Witchweed seriously impacts crop yield. Therefore, development of resistant varieties will have a great impact in developing countries, where stability of the food supply is of concern.

#### The Plant Genome Research Outreach Portal (PGROP)

Because of the great success of outreach activities conducted by NPGI project investigators, there are a large number of web-accessible resources for students and educators alike. Unfortunately, because of the diversity of web tools, it is often challenging to find available sites. Supplemental funding to the MaizeGDB project has facilitated the creation of the Plant Genome Research Outreach Portal (<http://www.plantgdb.org/outreach/>), which is designed to be a one-stop-shopping experience for students at all levels, their teachers, plant breeders, growers and extension specialists. The mission of this web site is to provide a centralized clearinghouse of Plant Genome “Outreach” programs and activities that is easily accessible to a wide-ranging audience. This web site has a variety of features that facilitates creation of sophisticated queries using pull-down menus. It allows users to easily add outreach resources, and should become the starting place for individuals looking for plant genomics materials for teaching and learning. It is also likely to engage researchers who stop at the PlantGDB website to use the research tools and who may not be aware of the exciting training programs that have grown out of plant genomics research.



## ***IV. New Projects Started in 2003***

In this section, examples of new projects supported in FY2003 are described. They represent the direction of plant genome research outlined in the new five-year plan for 2003-2005.

### Resources for the study of plant cell walls

The plant cell wall is an important biotechnological target because cell wall composition and structure constitute economically important traits ranging from tree mass and wood quality to plant feed nutritional quality. The advent of genomics is providing new insights, methods, and tools to study the biosynthesis, variation and chemical composition of cell walls. This is especially significant because the plant cell wall is extremely complex and difficult to study using traditional plant biochemistry methods. A collaborative effort between researchers at the Carnegie Institution of Washington and Oklahoma State University was initiated to use genomic and analytical chemical approaches to isolate, synthesize and characterize a set of recombinant enzymes (glycosyl hydrolases and polysaccharide lyases) that can break down specific components of plant cell walls. The activity and specificity of these polysaccharide-degrading enzymes will be determined; this information, along with the respective genetic clones and purified enzymes, will be made freely available to the science research community as a technical resource for the analysis of polysaccharides present in plant cell walls.

### Rice seed stock center

Several major efforts have recently taken hold in the research community to develop rice stocks for functional genomics research. These initiatives are designed to leverage the imminent completion of the rice genome sequence. In September 2002 and January 2003, a group of researchers held workshops to discuss how best to manage these resources for the community. Participants, who represented a broad spectrum of domestic and international plant genome researchers, quickly agreed that there was a need for a dedicated rice stock center that would enable access to rice genetic seed stocks. The discussions emphasized that many research communities would benefit from a rice resource center. Experience with *Arabidopsis*, *Drosophila* and other model organisms

whose genomes have been sequenced clearly demonstrates that access to resources through a resource center leads to research that previously was inconceivable.

Out of these discussions, the Dale Bumpers National Rice Research Center, an ARS/USDA facility in Stuttgart, Arkansas, has emerged as an ideally suited facility for rice genetic seed stocks. ARS has agreed to provide resources to establish the Center to become a premier rice seed stock center. The ARS-supported Maize Cooperative located in Champaign, Illinois will serve as an organizational model. The Maize



Cooperative has worked with the maize genomics research community to curate, catalogue, store, and distribute maize seed resources being produced in the course of research efforts under the NPGI. If the success of the Maize Cooperative is any indication, the Dale Bumpers National Rice Research Center will become an indispensable resource for the community, enabling any interested party to participate in plant genomics research.

#### *Medicago truncatula* sequencing project



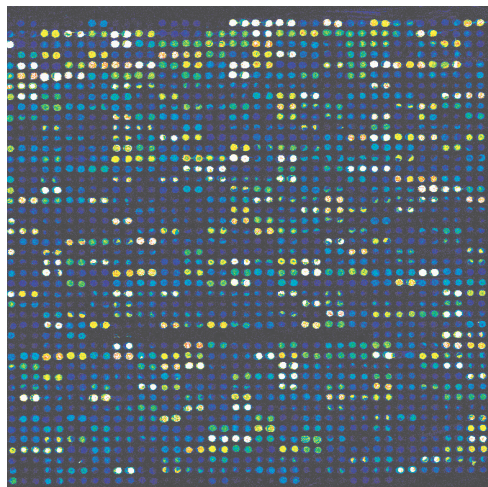
The more than 20,000 species of legumes together represent one of the two most important crop families in the world. The cultivated legumes include alfalfa, soybean, common bean, pea, lentil, lotus, and chickpea. Legumes are unique among cultivated plants in their ability to fix atmospheric nitrogen through symbiosis with bacteria known as *Rhizobia*. In part because of this unique symbiosis, nearly a third of all nutritional nitrogen comes from legumes, and they are consequently the single most important source of nutritional protein throughout the developing world. Legumes also synthesize secondary compounds with health promoting effects. Not surprisingly, legumes play a central role in nearly all cropping systems and are essential for secure and sustainable food production.

Development of genomic resources for the forage legume *Medicago truncatula* over the past 5 years has led to its emergence as a model system for all legumes. *Medicago* has a compact genome of approximately 470 million base pairs, facile Mendelian genetics, short generation time, relatively high transformation efficiency, and extensive collections of phenotypic mutants and naturally occurring genetic variants. Cytogenetic studies have revealed that the *Medicago* genome is organized into separate gene-rich and gene-poor regions. Indeed, gene density in the gene-rich portions of the *Medicago* genome is nearly as high as in *Arabidopsis*. This genome organization makes it possible to capture almost all of the genes by targeting just these gene-rich arms for sequencing.

A consortium of US researchers from the University of Minnesota, the University of Oklahoma, and the Institute for Genomics Research will partner with scientists from the European Union in an international effort to complete the sequence of the gene-rich portions of the eight *Medicago* chromosomes. There will be a variety of beneficiaries of the *Medicago* genome sequence including researchers studying basic biology and genomics, as well as breeders and growers of other legume crops, who will gain a reference genome sequence that includes the blueprint for the traits that make legumes such important crops worldwide. With the *Medicago* sequence in hand, detailed studies of legume-specific gene families, developmental processes, and biochemical pathways will be possible. In addition, gene cloning and marker selected breeding in other legume crops will be enhanced greatly by this 'roadmap' legume. Finally, comparison of the *Medicago* genome with *Arabidopsis* and rice will create enormous opportunities for comparative genomics, and should lead to an understanding of how symbiosis with microbes and the ability to fix nitrogen evolved.

### Community microarray centers for rice and maize

Two new projects are developing expression-profiling tools on a community-wide scale, building on experience gained from earlier, smaller-scale efforts in other plants. These new projects recognize that a group of scientists benefits when members of the community are provided microarrays and a hybridization service along with integrated tools for experimental design and data analysis. This organizational model permits scientists from institutions of all sizes and across the US to have access to technology that might otherwise be unavailable or too expensive. A team led by researchers at the University of California, Davis, and the University of Arizona, respectively, will produce rice and maize (corn) microarrays. Both projects will develop what should become the standard microarray for each plant and develop a database to house microarray expression data and provide statistical experimental design and analysis tools. Researchers will be able to purchase microarrays for use in their own facilities or use the service provided by the projects. All data resulting from the use of the microarrays will be deposited in a standard format that includes essential information about the experiment. It is anticipated that these community data will become a “browsable” resource that can be used to query genome-wide mRNA expression for genes or combinations of genes in response to a wide range of developmental and environmental situations. The benefits should include lower cost, reduced duplication of experimental treatments, and a more robust and reproducible public data set that will allow many individual investigators to benefit from genome-wide expression data in planning their experiments.



### TILLING in Maize

Mutants that have lost the function of a gene of interest are indispensable tools for understanding gene function. Existing programs for maize focus on making these mutants by a technique, transposon insertion, which tends to completely eliminate the gene function. Point mutations, which alter only one base in a gene, often result in subtle changes to the plant that are more informative in studies of how genes function and interact. In particular, a series of point mutations at every gene in the genome would be a powerful tool in these studies. Researchers at Purdue University and Iowa State University are establishing a community resource for discovery of point mutants in Maize. The ultimate goal of this population will be for use in TILLING, a screening technique that allows identification of plants with single base changes in any gene in the genome and web links to seed carrying the identified mutation.

TILLING is a technology developed at University of Washington with the initial support of a project in 2000. The proof of concept studies used Arabidopsis as the experimental system. Now, TILLING is being used in 16 plant species in 6 countries, including at the International Rice Research Institute (IRRI) in the Philippines. For technical details, go to <http://www.arabidopsis.org/help/quickstart.html>.



#### Nutritional genomics in sunflower

Sunflower (*Helianthus annuus L.*) oil is one of the most widely produced and consumed edible oils in the world. Researchers at Oregon State University are focusing on enhancing the nutritional characteristics of sunflower oil by manipulating natural genetic variants in wild and domesticated sunflowers and on developing tools and resources for manipulating nutritional traits in hybrid sunflower breeding programs. The objectives of this research are to identify genes for making the oil healthier by modifying the vitamin E and saturated fatty acid profiles. Sunflower naturally produces a broader range of vitamin E profiles than other crop plants. By focusing on understanding the genetic basis for the genetic variability in addition to producing genetically stable lines with diverse vitamin E profiles, manipulating genes that reduce saturated fat and increase monounsaturated or polyunsaturated fats will enhance the nutritional characteristics of sunflower oil.



#### Tomato fruit yield and quality

Fruit size and shape are two major factors determining yield, quality and consumer acceptability in many crops. Both are quantitatively inherited and have been difficult to approach with the tools of molecular biology. Researchers at Cornell University are conducting fundamental studies to identify, isolate and understand the molecular bases of the key loci underlying size and shape variation of tomato fruit. Results from the project will shed light on the nature and molecular basis of quantitative trait variation and contribute to a critical, but largely unexplored aspect of plant development: how ovaries are transformed from small reproductive organs into the large, conspicuous fruit that display the array of shapes and sizes that we associate with modern agriculture. In addition, this research may help to reconstruct events involved in the domestication of tomato and other fruit bearing crop species. This project is an extension of a previously funded project and takes full advantage of information and biological materials resulting from it.

#### Soybean disease resistance gene network

The genomes of many cultivated plants are polyploid, containing more than one copy of the genome. In the case of soybean, the genome underwent a complete duplication event about 9 million years ago to give rise to two nearly identical copies, and subsequent polyploidization events have occurred, including one 50,000 years ago. However, the changes in genome structure and organization that have occurred since the duplication are poorly understood, despite their potential impact on clusters of genes of agronomic importance such as for disease resistance. A project led by scientists at Indiana University will use comparative genomics approaches to examine the patterns of gene rearrangement on disease resistance genes. The duplicated resistance gene clusters from six different legume taxa are being sequenced, including two soybean cultivars, a close diploid (2N) relative of soybean, *Teramnus labialis*, and a recent polyploid, *Glycine tomentella*, that has twice as many chromosomes as soybean. Together, these sequences should yield a picture of the relative arrangements of these regions, including the portions that have remained unchanged and the regions that have diverged. It should be possible to use this information to gain a better picture of the impact of polyploidization on

disease resistance gene structure and function.

### PlantGDB – Informatics tools for the plant genome research community

Whole genome sequencing efforts for the *Arabidopsis* and rice genomes have captured the public imagination, and currently funded projects continue to produce an ever-increasing set of sequence resources that include Expressed Sequence Tags (ESTs), and genome survey sequences in the form of sequences flanking maize transposon insertion sites and sequences of gene-enriched genome fragments. However, this resource is only as useful as the tools developed to find and characterize the information it contains. The PlantGDB database (<http://www.plantgdb.org>), hosted at Iowa State University is developing tools to allow organization and interpretation of these sequence data. Two goals of the project are to estimate and characterize plant gene space, and the extent and conservation of alternative pre-mRNA splicing in plants. These objectives are being pursued by further development of algorithms and statistical methods for splice site recognition and gene structure prediction. As the database matures, researchers will be able to use the sequence information from a whole genome sequence such as that of rice in conjunction with EST and genome survey sequences from related cereals such as maize, sorghum, barley, and wheat, to build models of genes for other cereals such as millet where only a limited amount of sequence information is currently available.

## ***V. Plans for the Next Year***

The IWG outlined a broad series of objectives in the new five-year planning document, “National Plant Genome Initiative: 2003-2008”. Each participating agency plans to continue support of plant genome research based on the NPGI plan as appropriate for each agency’s mission.

The National Science Foundation (NSF) will continue to support activities covering all the six NPGI objectives, with special emphasis on: elucidation of the genome structure and function, functional genomics research, bioinformatics tool development, and the *Arabidopsis* 2010 Project. In addition, NSF’s agency-wide goals of “broadening participation”, “integration of research and education”, and “international research collaboration” are integrated into plant genome research projects supported by NSF.

The Department of Agriculture (USDA) plans to implement the NPGI plan through the National Research Initiative Competitive Grants Program (NRI) of the Cooperative State Research, Education and Extension Service (CSREES). Emphasis is on functional and translational genomics of cereals and basic genomics and bioinformatics of plants of agricultural and forestry relevance. USDA also provides essential support to the NPGI through the Agricultural Research Service (ARS). ARS will continue to provide long-term stable support for research databases such as MaizeGDB and Gramene, and for research resources such as Maize seed collections and rice seed collections.

The Department of Energy (DOE) expects that systems biology - an integration of genomics, computational, and imaging approaches and tools - will expand our ability to identify global networks of genes involved in the complex and specialized plant processes in growth, development and metabolism. Systems biology approaches seek to integrate all tools available to science to comprehend and ultimately predict biological processes. These approaches display great promise as researchers consider future demands for efficient and environmentally prudent renewal resource development, thus contributing to the mission of DOE and the function and translational genomics goals of the NPGI.

While the National Human Genome Research Institute (NHGRI) does not support plant genome research directly, activities supported by the NHGRI have contributed fundamental concepts and technologies in genomics, which the plant genome research community has taken advantage of and has built on. In April 2003, the NHGRI published a planning document outlining a vision for the future of genomics research at NIH. In the document, an emphasis is placed on continued development of research resources, new genomics technologies, computational biology, and training/education/societal impacts as an integral part of the future of genomics research. The plant genome research community will no doubt continue to benefit from the NIH investment in fundamental infrastructure for genomics research. Advances and developments in the human genome project will be communicated to the IWG through a member representing NIH, as well as involvement of NIH supported genomics researchers in the NPGI activities.

For the coming year, the IWG anticipates that submitted proposals will have increased focus on traits of

## *Plans for the Next Year*

---

economic importance as well as plant processes of fundamental importance. For example, such complex traits as nutritional quality or production of useful phytochemicals in plants are now possible to be studied using genomics approaches. Also these traits are becoming increasingly relevant and important, as the focus of plant biotechnology is moving toward addressing the interests of consumers. In addition, the IWG expects an increased interest in research collaboration between US researchers and scientists in developing countries. One of the joint activities being planned for implementation in the near future is a DOE/NSF/USAID/USDA joint program to establish just such research collaboration.

The IWG will continue to coordinate and provide oversight to the NPGI. More importantly, a robust line of communication has been firmly established at the program level, which contributes to interagency coordination of the NPGI. Representatives from the NPGI participating agencies attend each other's review panels and arrange joint review/funding as necessary. Any new planning workshops submitted from the community have been and will be supported jointly by appropriate agencies. As with any fast moving research area, it is expected that new and unexpected opportunities and challenges will emerge for the NPGI at any time. The IWG and NPGI participating agencies remain flexible and ready to take those opportunities as they arise.

## VI. Appendix

URLs for useful plant genome information resources

National Science and Technology Council homepage:

[http://www.ostp.gov/NSTC/html/NSTC\\_Home.html](http://www.ostp.gov/NSTC/html/NSTC_Home.html)

NSF Directorate for Biological Sciences homepage: <http://www.nsf.gov/bio>

USDA Research, Education and Economics homepage: <http://www.reeusda.gov/ree/>

DOE Office of Science homepage: <http://www.sc.doe.gov/>

National Human Genome Research Institute homepage:

<http://www.nhgri.nih.gov/>

National Plant Genome Initiative five-year plan: <http://www.ostp.gov/NSTC/html/npgi2003/index.htm>

NHGRI's vision for the future document: <http://www.nhgri.nih.gov/11006873>

NSF's FY2004-2005 program announcement for the Plant Genome Research Program:

<http://nsf.gov/pubs/2004/nsf04510/nsf04510.htm>

NRI's call for proposal for FY2004:

[http://www.reeusda.gov/1700/funding/04/pdf/rfa\\_nri\\_04.pdf](http://www.reeusda.gov/1700/funding/04/pdf/rfa_nri_04.pdf)

dbEST (EST database at the NCBI): <http://www.ncbi.nlm.nih.gov/dbEST/>

PlantGDB (General information about plant genome research): <http://www.plantgdb.org>

Plant Genome Databases (A collection of plant genome databases):

<http://www.hgmp.mrc.ac.uk/GenomeWeb/plant-gen-db.html>

PGROP (Resource for plant genome education, training and outreach): <http://www.plantgdb.org/outreach/>

MaizeGDB (Maize genome information resource): <http://www.maizegdb.org/>

Gramene (Cereal genome information resource): <http://www.gramene.org/>

TAIR (*Arabidopsis* information resource): <http://arabidopsis.org>

International Rice Genome Sequencing Project: <http://rgp.dna.affrc.go.jp/IRGSP/>

Maize sequencing project website: <http://www.maizegenome.org/>

*Medicago truncatula* Consortium: <http://www.medicago.org/>

# Acknowledgement

The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Jane Silverthorne (National Science Foundation), Robert Last (National Science Foundation), and Ed Kaleikau (U.S. Department of Agriculture) in preparation of this report. Victoria Bryan (National Science Foundation) provided assistance in production of the report.

The following individuals and organizations provided images used in this report:

Page 5 - Maize kernels; Dr. Ed Coe, MaizeGDB

Page 8 - Arabidopsis Greenhouse; Arabidopsis Biological Research Center, Ohio State University

Page 9 - *Solanum demissum* (late blight resistant wild potato); Paul Rangel, Baker Lab, University of California, Berkeley.

Page 10 (top left) - Wheat; Agricultural Research Service, USDA

Page 10 (bottom right) - The Public Seed Initiative; Dr. Molly Jahn, Cornell University

Page 12 - Gramene Data; Dr. Lincoln Stein, Cold Spring Harbor Laboratory

Page 14 - 2002 and 2003 ESPG Students; Dr. Thomas Brutnell, Boyce Thompson Institute

Page 16 - Dr. Neil Rutger at the Dale Bumpers National Rice Research Center; Dr. Kay Simmons, Agricultural Research Service, USDA

Page 17 - *Medicago truncatula*; Agricultural Research Service, USDA

Page 18 - Maize Endosperm microarray hybridized with polyA+ RNA from 14 DAP endosperm; Rangasamy Elumalai, David Galbraith and Vicki Chandler. Department of Plant Sciences, University of Arizona.

Page 19 - Sunflowers; Agricultural Research Service, USDA

## **Abstract**

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research program. The Interagency Working Group (IWG) on Plant Genomes provides coordination and oversight to the NPGI. The IWG published two long-range plans for the NPGI, the 1998-2002 plans in 1998 and the 2003-2008 plans in January 2003. As part of its activity, the IWG issues an annual progress report of the NPGI.

The current report describes highlights of recent progress in the field, with a primary focus on examples of accomplishments reported since January 2003. Research tools and research resources for plant genomics continue to accumulate. Data, information and other products of research are being shared freely and openly, allowing a broad community of scientists to apply genomics approaches to fundamental studies of plant biology. The same tools and resources are being applied to develop improved crops and new breeding strategies, as well. With the sequencing of the rice genome essentially complete, functional and translational genomics research in all cereal genomics are advancing at a rapid pace. A new international model-legume sequencing project promises to do the same for all legume genomics in a few years. There is every indication that plant genomics will continue to advance in the foreseeable future.

The report is also available on the NSTC Home Page at  
[http://www.ostp.gov/NSTC/html/NSTC\\_Home.html](http://www.ostp.gov/NSTC/html/NSTC_Home.html)





**EXECUTIVE OFFICE OF THE PRESIDENT**  
Office of Science and Techonolgy Policy  
Washington, D.C. 20502